



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Robust Stochastic Optimization For MISO Broadcasting Channel With Delayed CSIT and Limited Transmitting Antenna

Citation for published version:

Ratnarajah, T 2015, 'Robust Stochastic Optimization For MISO Broadcasting Channel With Delayed CSIT and Limited Transmitting Antenna', *IEEE Transactions on Vehicular Technology*.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

IEEE Transactions on Vehicular Technology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Robust Stochastic Optimization For MISO Broadcast Channel With Delayed CSIT and Limited Transmitting Antennas

Yi Luo, *Student Member, IEEE*, and Tharmalingam Ratnarajah, *Senior Member, IEEE*

Abstract—This work considers the K -user multiple-input-single-output (MISO) broadcast channel (BC) with delayed channel state information at transmitter (CSIT) where the base station (BS) has only *two* antennas. Based on the MAT¹ scheme, we propose a probabilistic-constraint optimization approach at the transmitter side to design the beamforming vectors by compromising between aligning interference and enhancing signal detectability. At the receiver side, minimal interference leakage method is proposed and the corresponding closed-form expressions for the interference suppression matrices are derived. For $K = 2$ case, the effective MIMO channel is derived and the closed-form expression of achievable data rate is obtained based on the proposed optimization method. Moreover, we show that the optimal scheme when $K = 3$ has the same number of effective channels as in the case of $K = 2$, which can be optimized by the probabilistic-constraint optimization method. For the case $K \geq 3$, we propose a new suboptimal scheme which achieves $\frac{4K}{3K-1}$ degrees-of-freedom and show that the new scheme can also be decomposed into the same number of effective channels as in the case of $K = 2$. Simulation result demonstrates interference leakage power versus the achievable data rate, where it can be seen that the achievable data rate increases slowly with the interference leakage until the optimal point and then decreases drastically. Simulation results also show that the proposed scheme outperforms the MAT scheme and dual signal-to-interference-noise ratio (SINR) scheme for various channel variances.

I. INTRODUCTION

As multiple antenna techniques such as multiple-input-single-output (MISO) and multiple-input-multiple-output (MIMO) become more and more sophisticated, the channel state information at the transmitter (CSIT) is becoming very crucial for achieving higher capacity with various techniques, such as water-filling algorithms, precoding techniques and interference alignment (IA) [2]. It should be noted that perfect and instantaneous CSIT is of great importance to all of the aforementioned techniques. Generally, CSIT acquisition is

realized through feedback from the receivers [3]. Specifically, before data transmission begins, the transmitter and receivers will go through a training period. During this period, the transmitter initiates by sending a specified number of pilot symbols, which are shared with the receivers. Receivers can estimate the channel and then convey the results back to the transmitter. In practice, the transmitter can hardly get perfect and instantaneous CSI for several reasons, one reason being feedback link not instantaneous. Moreover, the CSIT estimation error is proportional to the intensity of channel variations. A new scheme, known as the MAT scheme, was proposed in [1], which exploits only the delayed CSIT and successfully achieves degrees of freedom (DoF) higher than the schemes without any CSIT, e.g., Time Division Multiple Access (TDMA) that achieves *one* DoF. Furthermore, a closed-form expression of the DoF together with the optimal achievability scheme was presented in [1], when the number of antennas M is no less than the number of users. However, when $M < K$ and $K > 3$, the optimal achievability scheme and the DoF region have not been found yet.

A considerable amount of literature extend the work in [1] for the case when $M \geq K$ to time-correlated channel, which assume the current CSIT is correlated to the previous one and can be estimated. In [4], imperfect current CSIT with perfect delayed CSIT is considered in MISO broadcast channel (BC) and [5] has extended this idea further to the case of MIMO interference channel (IC) and BC. A study of mixed CSIT with imperfect current CSIT is provided in [6]. Moreover, DoF achieved with both imperfect current and imperfect delayed CSIT was investigated in [7] and [8]. Until now, theoretical studies on delayed CSIT have been introduced in multi-hop MIMO BC [9], and in networks concerning physical layer secrecy [10].

Recently, based on the available theoretical results, studies have been carried out to introduce optimization techniques to the schemes with delayed CSIT. In [11] and [12], the beamforming vectors are optimized in the general K -user MAT scheme, where minimum mean square error (MMSE) receiver and dual signal-to-interference-noise ratio (SINR) methods are realized. It should be noted that the dual SINR method is an iterative-free suboptimal algorithm. In [12], the MAT scheme is considered for $K = M$ and MMSE method is used to design the precoding vectors in the second phase, where the authors choose to maximize the lower bound of the sum rate by using Jensen's inequality to deal with the unknown channel entries. In [13], optimization of the MAT scheme is

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The project HIATUS acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission under FET-Open grant number HIATUS-265578. The work was also supported by the Seventh Framework Programme for Research of the European Commission under grant HARP-318489.

Y. Luo, and T. Ratnarajah are with the Institute for Digital Communications, School of Engineering, The University of Edinburgh, Alexander Graham Bell Building, Kings Buildings, Mayfield Road, Edinburgh, EH9 3JL, United Kingdom. (email: {y.luo; t.ratnarajah}@ed.ac.uk). Yi Luo is the corresponding author.

¹Named after M. A. Maddah-Ali and D. Tse [1].

considered which takes into account the training period and data length and shows the tradeoff between sum rate and the number of users.

In general, robust optimization techniques that deal with channel uncertainty can be categorized into two cases: worst-case method [14], [15] and stochastic method [16]. The relationship between the worst-case method and the stochastic method can be found in [14]. MIMO uplink with channel uncertainty is considered in [15], where stochastic optimization method is used. [17] considered maximization of the data rate under bounded uncertainty. A detail study on robust optimization with different class of uncertainty can be found in [18]. The most inspiring works are [19] and [14], where the authors consider single-input-single-output (SISO) network with Gaussian and bounded channel uncertainties.

In this work, we propose a robust stochastic optimization method based on the MAT scheme when $M \leq K$ (the equal only happens when $M = K = 2$), which is a practical assumption because there are number of cases in which the number of users is larger than the number of transmitting antennas. For example, base stations (BS) in cellular networks are usually equipped with 6 – 8 antennas while serving hundreds of users. The proposed work begins by obtaining the achievable sum rate to describe the relationship between the beamforming vectors and the channel vectors. Then, we formalize an optimization problem which maximizes the linear independence between two signal data streams while keeping high probability that the interference leakage power is under a certain level. The optimization method is based on the algorithm proposed in [15], [19], where real and imaginary parts of the channel are splitted to avoid complicated probability distribution. One of the major contributions of this work is that we take advantage of the channel distribution to formulate a probabilistic optimization problem, which brings robustness to the MAT scheme. Because the channel uncertainty of the proposed scheme is caused by incomplete knowledge of current channel that has Gaussian distribution, we constrain the interference leakage power with probabilistic method, which in turn realizes the robustness. We propose a suboptimal scheme for any K which achieves $\frac{4K}{3K-1}$ DoF and implement the robust optimization method in the scheme. Another contribution of this study is that the proposed optimization method is iterative-free, which reduces the signal processing complexity and delay of the network when fast fading is assumed. Based on the simulation results, we can witness that aligning the interference vectors weighs more than the case of increasing detectability of desired symbols. Meanwhile the proposed scheme achieves higher sum rate than the MAT scheme under correlated channel with various channel variances.

The remainder of this paper is organized as follows. Section II presents the system model employed for the proposed scheme and a brief introduction of the MAT scheme. Section III studies the case of $K = 2$, and presents the idea of the robust optimization method. Section IV extends the algorithm to the case when $K = 3$ and Section V presents the implementation of the new achievability scheme for any value of K . Simulations results are presented in Section VI. Concluding remarks are given in Section VII.

Notations: Throughout this paper, $(\cdot)^T$, $(\cdot)^H$ denote the transpose and conjugate transpose of a matrix respectively, while $\Pr(\cdot)$ stands for probability. The real part of complex number is written as $\text{Re}(\cdot)$ and imaginary part is $\text{Im}(\cdot)$. The operation $\|\cdot\|_F$ is the Frobenius norm and $\|\cdot\|$ is the Euclidean norm. The inverse function is represented by $(\cdot)^{-1}$ and \cap is the union operation. All the logarithmic functions used in this paper have base 2, i.e., $\log(\cdot) = \log_2(\cdot)$.

II. SYSTEM MODEL

We consider a K -user MISO downlink channel, where the BS is equipped with only *two* antennas and each user is equipped with *one* antenna. The transmitting power is limited to P with evenly distribution at the both antennas. A fast fading channel without time correlation is considered, hence no current CSIT can be estimated and all the channel entries are independent. Before transmitting the desired symbols, the BS sends pilots for channel estimation and therefore the receivers obtain perfect instantaneous CSI, i.e., the i -th receiver obtains the channel entries $\mathbf{h}_i(j) \in \mathbb{C}^{1 \times 2}$ which is associated with the j -th time slot. At the receiver side, the current perfect CSI is obtained and conveyed back to the BS at the end of each time slot. Moreover, as assumed in [6], [10], receivers have knowledge of all the beamforming vectors. In this way, the perfect CSI, $\mathbf{h}_i(j)$ is available at the BS within the k -th time slot, $k > j$, $\forall i = 1, 2, 4, \dots, K$. Since the proposed work follows from the MAT scheme, a brief introduction and analysis of the MAT scheme is given in the next subsection for the case of $K = 2$ and $K = 3$.

A. MAT scheme when $K = 2$

The scheme consists of *three* time slots and two transmitted symbol vectors $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{C}^{1 \times 2}$, each of which contains two private symbols. The BS sends *two* private symbol vectors to each receiver using the first two time slots, and then sends the sum of interference generated in the first two time slots with the obtained delayed CSIT as the symbols for the third time slot. If each row represents one time slot, the symbols received at the i -th receiver are given by

$$\mathbf{y}_i = \sqrt{\frac{P}{2}} \underbrace{\begin{bmatrix} \mathbf{h}_i(1) \\ \mathbf{0} \\ h_{i,1}(3)\mathbf{h}_2(1) \end{bmatrix}}_{\mathbf{H}_{i1}} \mathbf{s}_1 + \sqrt{\frac{P}{2}} \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{h}_i(2) \\ h_{i,1}(3)\mathbf{h}_1(2) \end{bmatrix}}_{\mathbf{H}_{i2}} \mathbf{s}_2 + \mathbf{n}_i, \quad (1)$$

where $\mathbf{h}_i(j)$ is the channel to Rx- i at the j -th time slot and $h_{i,1}(3)$ is the channel from the first antenna at the BS to Rx- i at the third time slot and $\mathbf{n}_i = [n_i(1) \ n_i(2) \ n_i(3)]^T$ is the independent additive white noise at each time slot with unit power. We assume all the entries of the channel vectors are independent and identical distributed (i.i.d.) as complex Gaussian variables. It can be shown that \mathbf{H}_{11} and \mathbf{H}_{22} have rank *two* almost surely because the channel vectors are independent at the two receivers. Similarly, we can show that \mathbf{H}_{12} and \mathbf{H}_{21} are rank *one* matrices. With this property, receivers gain two independent observations of their two desired symbols

and they can eliminate the interference simply by making a subtraction. Specifically, the corresponding interference suppression matrices to decode the private symbols at Rx-1 and Rx-2 should be given by

$$\bar{\mathbf{U}}_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -h_{1,1}(3) & 1 \end{bmatrix}, \quad \bar{\mathbf{U}}_2^T = \begin{bmatrix} 0 & 1 & 0 \\ -h_{2,1}(3) & 0 & 1 \end{bmatrix}.$$

B. MAT scheme when $K = 3$

The optimal scheme for $K = 3$ has been given in [1] which achieves the outer bound, i.e., $3/2$ total DoF. The optimal scheme is divided into 3 phases and 8 time slots. For each receiver, there are 2 desired symbol vectors, namely \mathbf{s}_i and \mathbf{s}'_i , where $i = 1, 2, 3$. The first phase contains three time slots, where the BS sends private symbol vectors. Specifically, the BS sends $\mathbf{s}_1 + \mathbf{s}_2$ in the first time slot, $\mathbf{s}'_1 + \mathbf{s}_3$ in the second time slot and $\mathbf{s}'_2 + \mathbf{s}'_3$ in the third time slot. In each time slot of the second phase, the BS reconstructs the interference generated at the corresponding time slot in the first phase. Therefore, there are 3 other time slots in the second phase. The third phase has two time slots, where two independent linear combinations of the symbol vectors reconstructed as interference received in the second phase, i.e. linear combinations of $y_1(4), y_2(5), y_3(6)$, are transmitted, where $y_i(t) \in \mathbb{C}$ is the symbol received at Rx- i at time slot t . A complete collection of the transmitted symbols in the scheme are given in Table I.

Transmissions in the third phase only uses one antenna and the symbols can be written as

$$\begin{aligned} x_i = & \underbrace{a_i \mathbf{h}_3(4) \begin{bmatrix} \mathbf{h}_2(1)\mathbf{s}_1 \\ \mathbf{h}_1(1)\mathbf{s}_2 \end{bmatrix}}_{y_3(4)} + \underbrace{b_i \mathbf{h}_2(5) \begin{bmatrix} \mathbf{h}_3(2)\mathbf{s}'_1 \\ \mathbf{h}_1(2)\mathbf{s}_3 \end{bmatrix}}_{y_2(5)} \\ & + \underbrace{c_i \mathbf{h}_1(6) \begin{bmatrix} \mathbf{h}_3(3)\mathbf{s}'_2 \\ \mathbf{h}_1(3)\mathbf{s}'_3 \end{bmatrix}}_{y_1(6)}, \end{aligned} \quad (2)$$

where $i = 1, 2$ and a_i, b_i, c_i are constants chosen randomly by the BS but shared with the receivers. When the receivers obtain symbols from all the 8 time slots, they will be able to decode the desired symbols using backward decoding. Totally 12 symbols are sent within 8 time slots, which means $3/2$ DoF is achieved.

C. General number of users when $K > 3$

Unlike the case when $K \leq 3$, where scheme achieving the outer bound proposed in [1] is available, in cases when $K > 3$, it is still an open problem whether the outer bound given in [1] is tight or not. Therefore, there is no available scheme we can generalize like the MAT scheme for $K \leq 3$ and we propose a suboptimal scheme which extended from the case of $K = 3$. Moreover, we utilize our robust stochastic optimization method to maximize the total achievable data rate in the general K -user case.

III. ROBUST STOCHASTIC OPTIMIZATION FOR $K = 2$ CASE

In this section, we firstly generalize the MAT scheme in order to design flexible beamforming vectors. Next, we maximize the achievable sum rate with consideration of $h_{i1}(3)$, $i = 1, 2$, which is unknown to the BS. Unlike [11], where $h_{i1}(3)$ is handled by Jensen's inequality, we exploit the statistical CSI and make probabilistic constraint, which brings robustness to the algorithm.

After obtaining the delayed CSIT, instead of reconstructing the interference exactly the way they are, i.e., $\mathbf{h}_2(1)\mathbf{s}_1 + \mathbf{h}_1(2)\mathbf{s}_2$, the BS constructs the symbols vector to be sent at the third time slot as $\mathbf{x}(3) = \mathbf{w}_1\mathbf{s}_1 + \mathbf{w}_2\mathbf{s}_2$, where $\mathbf{x}(i)$ is the vector to be sent at time slot i . Although the flexibility given to the beamforming vector is based on the work [12], we further utilize the statistical information of the unknown current CSIT when investigating the achievable sum rate. This consideration not only brings robustness through the beamforming vectors, but also provides flexible interference suppression matrix. The signals received during all the three time slots can be written as

$$\begin{aligned} \mathbf{y}_i = & \begin{bmatrix} y_i(1) \\ y_i(2) \\ y_i(3) \end{bmatrix} = \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{h}_i(1) \\ \mathbf{0} \\ h_{i1}(3)\mathbf{w}_1 \end{bmatrix} \mathbf{s}_1 \\ & + \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{0} \\ \mathbf{h}_i(2) \\ h_{i1}(3)\mathbf{w}_2 \end{bmatrix} \mathbf{s}_2 + \mathbf{n}_i, i, j = 1, 2, i \neq j. \end{aligned} \quad (3)$$

where $\mathbf{n}_i = [n_i(1) \ n_i(2) \ n_i(3)]^T$ denotes the additive white noise and $\mathbb{E}[\mathbf{n}_i^H \mathbf{n}_i] = \mathbf{I}^{(3)}$, where $\mathbf{I}^{(3)}$ is the 3-dimensional identity matrix. Let us consider the receiver Rx-1, in the first time slot, there is only the desired signal and in the second time slot, there is only interference. In order to decode the desired symbol vector, Rx-1 subtracts $y_1(2) = \mathbf{h}_1(2)\mathbf{s}_2$ from $y_1(3)$. Therefore, Rx-1 have two independent linear combinations of u_1, v_1 with reduced interference. Correspondingly, Rx-1 can construct the interference suppression matrix which minimize the interference leakage power is given by

$$\mathbf{U}_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -x_1 & 1 \end{bmatrix}, \quad \mathbf{U}_2^T = \begin{bmatrix} 0 & 1 & 0 \\ -x_2 & 0 & 1 \end{bmatrix}, \quad (4)$$

where $x_i, i = 1, 2$ is the complex scalars that are to be designed by the receivers. After multiplying with the interference suppression matrix, the received signals are given by

$$\begin{aligned} \bar{\mathbf{y}}_1 = & \mathbf{U}_1^T \mathbf{y}_1 = \sqrt{\frac{P}{2}} \underbrace{\begin{bmatrix} \mathbf{h}_1(1) \\ h_{1,1}(3)\mathbf{w}_1 \end{bmatrix}}_{\mathbf{H}_{11}} \mathbf{s}_1 \\ & + \sqrt{\frac{P}{2}} \underbrace{\begin{bmatrix} \mathbf{0} \\ h_{1,1}(3)\mathbf{w}_2 - x_1\mathbf{h}_1(2) \end{bmatrix}}_{\mathbf{H}_{12}} \mathbf{s}_2 + \begin{bmatrix} n_1(1) \\ n_1(3) - x_1 n_1(2) \end{bmatrix}, \end{aligned} \quad (5)$$

TABLE I
SYMBOLS RECEIVED IN ALL THE TIME SLOTS

		first time slot	second time slot	third time slot
Phase 1	Rx-1	$\mathbf{h}_1(1)(\mathbf{s}_1 + \mathbf{s}_2)$	$\mathbf{h}_1(2)(\mathbf{s}'_1 + \mathbf{s}_3)$	$\mathbf{h}_1(3)(\mathbf{s}'_2 + \mathbf{s}'_3)$
	Rx-2	$\mathbf{h}_2(1)(\mathbf{s}_1 + \mathbf{s}_2)$	$\mathbf{h}_2(2)(\mathbf{s}'_1 + \mathbf{s}_3)$	$\mathbf{h}_2(3)(\mathbf{s}'_2 + \mathbf{s}'_3)$
	Rx-3	$\mathbf{h}_3(1)(\mathbf{s}_1 + \mathbf{s}_2)$	$\mathbf{h}_3(2)(\mathbf{s}'_1 + \mathbf{s}_3)$	$\mathbf{h}_3(3)(\mathbf{s}'_2 + \mathbf{s}'_3)$
Phase 2	Rx-1	$\mathbf{h}_1(4) [\mathbf{h}_2(1)\mathbf{s}_1 \ \mathbf{h}_1(1)\mathbf{s}_2]^T$	$\mathbf{h}_1(5) [\mathbf{h}_3(2)\mathbf{s}'_1 \ \mathbf{h}_1(2)\mathbf{s}_3]^T$	$\mathbf{h}_1(6) [\mathbf{h}_3(3)\mathbf{s}'_2 \ \mathbf{h}_1(3)\mathbf{s}'_3]^T$
	Rx-2	$\mathbf{h}_2(4) [\mathbf{h}_2(1)\mathbf{s}_1 \ \mathbf{h}_1(1)\mathbf{s}_2]^T$	$\mathbf{h}_2(5) [\mathbf{h}_3(2)\mathbf{s}'_1 \ \mathbf{h}_1(2)\mathbf{s}_3]^T$	$\mathbf{h}_2(6) [\mathbf{h}_3(3)\mathbf{s}'_2 \ \mathbf{h}_1(3)\mathbf{s}'_3]^T$
	Rx-3	$\mathbf{h}_3(4) [\mathbf{h}_2(1)\mathbf{s}_1 \ \mathbf{h}_1(1)\mathbf{s}_2]^T$	$\mathbf{h}_3(5) [\mathbf{h}_3(2)\mathbf{s}'_1 \ \mathbf{h}_1(2)\mathbf{s}_3]^T$	$\mathbf{h}_3(6) [\mathbf{h}_3(3)\mathbf{s}'_2 \ \mathbf{h}_1(3)\mathbf{s}'_3]^T$
Phase 3	Rx-1	$h_{1,1}(7)x_1$	$h_{1,1}(8)x_2$	Empty
	Rx-2	$h_{2,1}(7)x_1$	$h_{2,1}(8)x_2$	
	Rx-3	$h_{3,1}(7)x_1$	$h_{3,1}(8)x_2$	

$$\begin{aligned} \bar{\mathbf{y}}_2 = \mathbf{U}_2^T \mathbf{y}_2 = \sqrt{\frac{P}{2}} \underbrace{\begin{bmatrix} \mathbf{h}_2(2) \\ h_{2,1}(3)\mathbf{w}_2 \end{bmatrix}}_{\mathbf{H}_{22}} \mathbf{s}_2 \\ + \sqrt{\frac{P}{2}} \underbrace{\begin{bmatrix} \mathbf{0} \\ h_{2,1}(3)\mathbf{w}_1 - x_2\mathbf{h}_2(1) \end{bmatrix}}_{\mathbf{H}_{21}} \mathbf{s}_1 + \begin{bmatrix} n_2(2) \\ n_2(3) - x_2n_2(1) \end{bmatrix}, \end{aligned} \quad (6)$$

where the effective MIMO channels have *zero* interference in the first data stream and $h_{1,1}(3)\mathbf{w}_2 - x_1\mathbf{h}_1(2)$ in the second one. To help better clarify our idea in the subsequent sections, we define in the system model (5) and (6) as *e-channels*, where channel for one of the data streams has uncertain variable and the interference is difference of two vectors. If we let $x_1 = h_{1,1}(3)$, $x_2 = h_{2,1}(3)$ and $\mathbf{w}_1 = \mathbf{h}_2(1)$, $\mathbf{w}_2 = \mathbf{h}_1(2)$, then the scheme reduces to the MAT scheme. Specifically, there exists a tradeoff between reducing the interference leakage power (aligning interference in two time slots) and increasing the detectability of the desired symbols (linear independence between two signal data streams), where the MAT scheme is on one side of the tradeoff and the proposed work is trying to strike a balance in order to maximize the achievable sum rate.

Theorem 1: The total achievable sum rate achieved in both receivers can be approximated by the following equation:

$$\begin{aligned} I(\mathbf{s}_1; \mathbf{y}_1) + I(\mathbf{s}_2; \mathbf{y}_2) \approx \log \left(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2 \right) + \log \left(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2 \right) \\ + \log \left(\frac{\frac{P}{2} |h_{1,1}(3)|^2 + \frac{P^2}{4} |h_{1,1}(3)|^2 (\|\mathbf{h}_1(1)\|^2 - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H)}{(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2) \Theta_2 \hat{x}_2^2} \right) \\ + \log \left(\frac{\frac{P}{2} |h_{2,1}(3)|^2 + \frac{P^2}{4} |h_{2,1}(3)|^2 (\|\mathbf{h}_2(2)\|^2 - \mathbf{w}_2 \mathbf{h}_2^H(2) \mathbf{h}_2(2) \mathbf{w}_2^H)}{(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2) \Theta_1 \hat{x}_1^2} \right). \end{aligned}$$

where

$$\begin{aligned} \Theta_1 &= 1 + \frac{P}{2} \|h_{1,1}(3)\mathbf{w}_2 - x_1\mathbf{h}_1(2)\|^2, \\ \Theta_2 &= 1 + \frac{P}{2} \|h_{2,1}(3)\mathbf{w}_1 - x_2\mathbf{h}_2(1)\|^2, \\ \hat{x}_i^2 &= \max\{1, |x_i|^2\}. \end{aligned}$$

Proof: See Appendix A.

In order to simplify the remaining description, we define the numerator that contains only \mathbf{w}_i as $\Phi_i(\mathbf{w}_i)$ and the denominator that contains only \mathbf{w}_i as $\Psi_i(\mathbf{w}_i)$. Based on the Theorem 1, we have

$$\begin{aligned} \Phi_1(\mathbf{w}_1) &= \frac{P}{2} |h_{1,1}(3)|^2 \\ &+ \frac{P^2}{4} |h_{1,1}(3)|^2 (\|\mathbf{h}_1(1)\|^2 - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H), \end{aligned} \quad (8a)$$

$$\Psi_1(\mathbf{w}_1) = \left(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2 \right) \Theta_2 \hat{x}_2^2, \quad (8b)$$

$$\begin{aligned} \Phi_2(\mathbf{w}_2) &= \frac{P}{2} |h_{2,1}(3)|^2 \\ &+ \frac{P^2}{4} |h_{2,1}(3)|^2 (\|\mathbf{h}_2(2)\|^2 - \mathbf{w}_2 \mathbf{h}_2^H(2) \mathbf{h}_2(2) \mathbf{w}_2^H), \end{aligned} \quad (8c)$$

$$\Psi_2(\mathbf{w}_2) = \left(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2 \right) \Theta_1 \hat{x}_1^2, \quad (8d)$$

where it can be observed that in each logarithmic function, both numerator and denominator contain the same beamforming vector, i.e., either \mathbf{w}_1 or \mathbf{w}_2 . Hence, in order to maximize the sum rate, we can equivalently find each optimal beamforming vector and thus optimizing each logarithmic function independently.

For the rest of the derivations, we only maximize the first logarithmic function because all the channel entries are i.i.d, the second one can be solved in a similar way. As the non-convexity of the logarithmic functions in terms of \mathbf{w}_i , it is not possible to maximize the sum rate directly. Thus $\|\mathbf{h}_1(1)\|^2 - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H$ is maximized and Θ_2 is minimized alternatively. The former item can be transformed with the following equation:

$$\begin{aligned} &\|\mathbf{h}_1(1)\|^2 - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H \\ &= \|\mathbf{h}_1(1)\|^2 \mathbf{w}_1 \mathbf{I} \mathbf{w}_1^H - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H \\ &= \mathbf{w}_1 (\|\mathbf{h}_1(1)\|^2 \mathbf{I} - \mathbf{h}_1^H(1) \mathbf{h}_1(1)) \mathbf{w}_1^H \\ &= \mathbf{w}_1 (\mathbf{h}_1^H(1))^\perp \mathbf{h}_1^\perp(1) \mathbf{w}_1^H, \end{aligned} \quad (9)$$

which is based on

$$(\mathbf{h}_1^H(1))^\perp \mathbf{h}_1^\perp(1) + \mathbf{h}_1^H(1)\mathbf{h}_1(1) = \|\mathbf{h}_1(1)\|^2 \mathbf{I}. \quad (10)$$

Therefore, the objective function of the maximization problem given as

$$\max_{\mathbf{w}_1} \text{trace} \left(\mathbf{w}_1 (\mathbf{h}_1^H(1))^\perp \mathbf{h}_1^\perp(1) \mathbf{w}_1^H \right). \quad (11)$$

In addition, because the uncertainty of the e-channel is caused by the absence of knowledge of the current CSI, which is Gaussian distributed, it is useful to exploit stochastic methods to handle the uncertainty. Specifically, we constrain the probability that interference leakage power Θ_2 satisfies the following:

$$\text{s.t. } \Pr(\|\mathbf{h}_{2,1}(3)\mathbf{w}_1 - x_2\mathbf{h}_2(1)\|^2 \leq \bar{\alpha}) \geq \beta, \quad (12)$$

where $0 \leq \beta \leq 1$ and $\bar{\alpha}$ is parameter to be chosen according to the requirement of interference leakage. Because of the difficulties to handle (12) directly, we need to make the following transformation. By regarding x_2 as any complex number that is to be chosen, the norm can be interpreted as the distance between any point in the line at the direction of $\mathbf{h}_2(1)$ and to the point $\mathbf{h}_{2,1}(3)\mathbf{w}_1$. According to the property of complex space, cosine of the angle τ between \mathbf{w}_1 and $\mathbf{h}_2(1)$ is $\cos(\tau) = \frac{\text{Re}(\mathbf{w}_1 \mathbf{h}_2^H(1))}{\|\mathbf{w}_1\| \cdot \|\mathbf{h}_2(1)\|}$. Intuitively, the nearest distance from the direction $\mathbf{h}_2(1)$ to the point $\mathbf{h}_{2,1}(3)\mathbf{w}_1$ is $\sin(\tau)|\mathbf{h}_{2,1}(3)|\|\mathbf{w}_1\|$. An important notation of the latter property is that because \mathbf{w}_1 has unit norm and $\mathbf{h}_2(1)$ can not be chosen, the distance is proportional to the norm of $\mathbf{h}_{2,1}(3)$, which explains the robustness of the proposed method. Meanwhile, when \mathbf{w}_1 approaches the direction of $\mathbf{h}_2(1)$ in complex space, the distance reduces and the distance equals to *zero* only when \mathbf{w}_1 and $\mathbf{h}_2(1)$ are in the same direction. When \mathbf{w}_1 approaches $\mathbf{h}_2(1)$ it can be equivalently expressed as \mathbf{w}_1 being orthogonal to $(\mathbf{h}_2^H(1))^\perp$ in complex space. Therefore, the probabilistic constraint can be rewritten as

$$\Pr(|\mathbf{h}_{2,1}(3)|^2 \text{trace}(\mathbf{w}_1 (\mathbf{h}_2^H(1))^\perp \mathbf{h}_2^\perp(1) \mathbf{w}_1^H) \leq \alpha) \geq \beta. \quad (13)$$

where α is also a parameter to be chosen by the BS. In this way, the overall optimization problem is formalized as

$$\max_{\mathbf{w}_1} \text{trace}(\mathbf{w}_1 (\mathbf{h}_1^H(1))^\perp \mathbf{h}_1^\perp(1) \mathbf{w}_1^H) \quad (14a)$$

$$\text{s.t. } \Pr(|\mathbf{h}_{2,1}(3)|^2 \text{trace}(\mathbf{w}_1 (\mathbf{h}_2^H(1))^\perp \mathbf{h}_2^\perp(1) \mathbf{w}_1^H) \leq \alpha) \geq \beta \quad (14b)$$

$$\|\mathbf{w}_1\|^2 = 1 \quad (14c)$$

Because all the channel entries are i.i.d., the optimization problem regarding to \mathbf{w}_2 can be found in a similar way.

Theorem 2: The probabilistic constraint in the optimization problem can be translated into the following inequality

$$\text{trace}(\mathbf{w}_1 (\mathbf{h}_2^H(1))^\perp \mathbf{h}_2^\perp(1) \mathbf{w}_1^H) \leq \frac{\gamma}{4\sigma_h^2}, \quad (15)$$

where

$$\gamma \triangleq \frac{\alpha}{[\text{erf}^{-1}(\sqrt{\beta})]^2}, \quad (16)$$

is the only parameter that should be chosen by transmitters

according to the channel variance.

Proof: See Appendix B.

Remark: Unfortunately, the closed-form expression of the optimal γ that maximizes the achievable sum rate can not be obtained. However, a numerical approach will be illustrated in the simulation section which represents the basis to select γ .

Based on Theorem 2, the optimization problem of \mathbf{w}_1 can be equivalently written as

$$\max_{\mathbf{w}_1} \text{trace}(\mathbf{w}_1 (\mathbf{h}_1^H(1))^\perp \mathbf{h}_1^\perp(1) \mathbf{w}_1^H) \quad (17a)$$

$$\text{s.t. } \text{trace}(\mathbf{w}_1 (\mathbf{h}_2^H(1))^\perp \mathbf{h}_2^\perp(1) \mathbf{w}_1^H) \leq \frac{\gamma}{4\sigma_h^2} \quad (17b)$$

$$\|\mathbf{w}_1\|^2 = 1 \quad (17c)$$

Because the objective function is non-convex, the optimization problem should be translated into the following one,

$$\max_{\mathbf{W}_1} \text{trace}(\mathbf{h}_1^\perp(1) \mathbf{W}_1 (\mathbf{h}_1^H(1))^\perp) \quad (18a)$$

$$\text{s.t. } \text{trace}(\mathbf{h}_2^\perp(1) \mathbf{W}_1 (\mathbf{h}_2^H(1))^\perp) \leq \frac{\gamma}{4\sigma_h^2} \quad (18b)$$

$$\text{trace}(\mathbf{W}_1) = 1 \quad (18c)$$

$$\mathbf{W}_1 \succcurlyeq \mathbf{0} \quad (18d)$$

where $\mathbf{W}_i \triangleq \mathbf{w}_i \mathbf{w}_i^H, \forall i$, $\mathbf{W}_i \succcurlyeq \mathbf{0}$ means that \mathbf{W}_i is positive semi-definite matrices. As proved in [20], \mathbf{W}_1 will surely have rank one. Therefore, the singular vector of the solution of (18a)-(18d) is equivalent to the solution of (17a)-(17c). As \mathbf{W}_1 is positive semi-definite, the optimization problem (18a)-(18d) forms a semi-definite programming (SDP) problem and can be efficiently solved by optimization tools such as CVX tool in Matlab.

The two receivers can use the following optimization method to design the receive signal matrices $\mathbf{U}_1, \mathbf{U}_2$ so that the interference leakage power is minimized, i.e.,

$$\min_{x_1 \in \mathbb{C}} \|\mathbf{H}_{12}\|_F^2, \quad \min_{x_2 \in \mathbb{C}} \|\mathbf{H}_{21}\|_F^2, \quad (19)$$

which have closed-form solution for the inside variables as

$$x_1 = \mathbf{h}_{1,1}(3) \mathbf{w}_2 \mathbf{h}_1^H(2) (\mathbf{h}_1(2) \mathbf{h}_1^H(2))^{-1} \quad (20a)$$

$$x_2 = \mathbf{h}_{2,1}(3) \mathbf{w}_1 \mathbf{h}_2^H(1) (\mathbf{h}_2(1) \mathbf{h}_2^H(1))^{-1} \quad (20b)$$

By this stage, the iterative-free algorithm is completed.

IV. ROBUST STOCHASTIC OPTIMIZATION FOR $K = 3$ CASE

For the case of $K = 3$, the MAT scheme in [1] demonstrates the optimal scheme which achieves the outer bound. In this section, we will show that the optimal MAT scheme, which takes 3 phases and 8 time slots, can be extended by generalizing the beamforming vector and decomposed into 3 independent e-channels. Therefore, we can utilize the robust stochastic optimization method to maximize the achievable sum rate.

It can be seen from Table I that no beamforming technique is required in the first phase as well as the third phase, where symbols that are received in the second phase are sent. In the second phase, the beamforming technique is exploited when the BS sends the reconstructed interference seen by receivers

at the corresponding time slot in the first phase. In the MAT scheme, the BS reconstructed symbol vectors in the second phase can be written as

$$\begin{aligned} \mathbf{x}_{MAT}(4) &= \begin{bmatrix} \mathbf{h}_2(1)\mathbf{s}_1 \\ \mathbf{h}_1(1)\mathbf{s}_2 \end{bmatrix}, \mathbf{x}_{MAT}(5) = \begin{bmatrix} \mathbf{h}_3(2)\mathbf{s}'_1 \\ \mathbf{h}_1(2)\mathbf{s}_3 \end{bmatrix}, \\ \mathbf{x}_{MAT}(6) &= \begin{bmatrix} \mathbf{h}_3(3)\mathbf{s}'_2 \\ \mathbf{h}_1(3)\mathbf{s}'_3 \end{bmatrix}. \end{aligned} \quad (21)$$

Instead of reconstructing the interference exactly the way they are in (21), we construct the symbol vectors to be sent as a flexible combination of desired symbols. More specifically, we realize the tradeoff between reducing more interference and generating more desired signal detectability in the new constructed symbols. The first step is to generalize the beamforming vectors as follows

$$\mathbf{x}(4) = \begin{bmatrix} \mathbf{w}_1\mathbf{s}_1 \\ \mathbf{w}_2\mathbf{s}_2 \end{bmatrix}, \quad \mathbf{x}(5) = \begin{bmatrix} \mathbf{w}'_1\mathbf{s}'_1 \\ \mathbf{w}_3\mathbf{s}_3 \end{bmatrix}, \quad \mathbf{x}(6) = \begin{bmatrix} \mathbf{w}'_2\mathbf{s}'_2 \\ \mathbf{w}'_3\mathbf{s}'_3 \end{bmatrix} \quad (22)$$

In this way, the signals received in the first and second phases at Rx- i can be rewritten in a much straightforward manner as in (23), where only the symbols received in the first two phases are listed so as to better concentrate on the proposed algorithm. The reason behind is that the third phase serves only to help receivers decoding each term in the second phase. Specifically, in the third phase of the proposed scheme, the BS sends two independent linear combinations of $y_3(4)$, $y_2(5)$ and $y_1(6)$. Hence, aided with the knowledge of $y_1(6)$ that is received at 6-th time slot by Rx-1, $y_3(4)$ and $y_2(5)$ can successfully obtain as $y_3(4) = \mathbf{h}_3(4)[\mathbf{w}_1\mathbf{s}_1 \ \mathbf{w}_2\mathbf{s}_2]^T$ and $y_2(5) = \mathbf{h}_2(5)[\mathbf{w}'_1\mathbf{s}'_1 \ \mathbf{w}_3\mathbf{s}_3]^T$. By making use of the observations $y_1(4)$ and $y_1(5)$, we obtain the following

$$\begin{aligned} \begin{bmatrix} y_1(4) \\ y_3(4) \end{bmatrix} &= \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{h}_1(4) \\ \mathbf{h}_3(4) \end{bmatrix} \begin{bmatrix} \mathbf{w}_1\mathbf{s}_1 \\ \mathbf{w}_2\mathbf{s}_2 \end{bmatrix}, \\ \begin{bmatrix} y_1(5) \\ y_2(5) \end{bmatrix} &= \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{h}_1(5) \\ \mathbf{h}_2(5) \end{bmatrix} \begin{bmatrix} \mathbf{w}'_1\mathbf{s}'_1 \\ \mathbf{w}_3\mathbf{s}_3 \end{bmatrix}, \end{aligned} \quad (24)$$

with which Rx-1 can successfully decode $\mathbf{w}_2\mathbf{s}_2$ and $\mathbf{w}_3\mathbf{s}_3$. The next task is to find two linear combinations of symbols in \mathbf{s}_1 and \mathbf{s}'_1 respectively which enables us to decode all the desired symbols. With all the received signals and interference related to \mathbf{s}_1 and \mathbf{s}'_1 in hand, Rx-1 can successfully obtain (25), where x_1, x'_1 are complex scalars chosen by Rx-1. We can see from (25) that Rx-1 can independently decode \mathbf{s}_1 and \mathbf{s}'_1 , where we have two e-channels to decode each of them. More importantly, the effective MIMO channel to decode the private symbols is the same as the effective MIMO channel in the 2-user case. This means that, similar to (44), the achievable sum rate of \mathbf{s}_1 can be written as

$$I(\mathbf{s}_1; y_1(1), y_1(4), \mathbf{w}_2\mathbf{s}_2) = \mathcal{C} + \log \left(\frac{\Phi_1(\mathbf{w}_1)}{\Psi_2(\mathbf{w}_2)} \right), \quad (26)$$

where \mathcal{C} stands for all the terms independent with \mathbf{w}_i and $\Phi_i(\mathbf{w}_i)$, $\Psi_i(\mathbf{w}_i)$ represent functions related to the precoding

vector \mathbf{w}_i and independent of any another $\mathbf{w}_j, j \neq i$. Because the effective MIMO channel to decode the private symbols is the same as that in the 2-user case, the specific realizations of $\Phi_i(\cdot)$ and $\Psi_i(\cdot)$ can be found with similar method as obtaining (8a)-(8d). The sum of achievable sum rate to decode all the desired symbols is given by

$$\begin{aligned} I_\Sigma &= \mathcal{C} + \log \left(\frac{\Phi_1(\mathbf{w}_1)}{\Psi_2(\mathbf{w}_2)} \right) + \log \left(\frac{\Phi'_1(\mathbf{w}'_1)}{\Psi_3(\mathbf{w}_3)} \right) + \log \left(\frac{\Phi_2(\mathbf{w}_2)}{\Psi_1(\mathbf{w}_1)} \right) \\ &\quad + \log \left(\frac{\Phi'_2(\mathbf{w}'_2)}{\Psi'_3(\mathbf{w}'_3)} \right) + \log \left(\frac{\Phi_3(\mathbf{w}_3)}{\Psi'_1(\mathbf{w}'_1)} \right) + \log \left(\frac{\Phi'_3(\mathbf{w}'_3)}{\Psi'_2(\mathbf{w}'_2)} \right) \\ &= \mathcal{C} + \log \left(\frac{\Phi_1(\mathbf{w}_1)}{\Psi_1(\mathbf{w}_1)} \right) + \log \left(\frac{\Phi'_1(\mathbf{w}'_1)}{\Psi'_1(\mathbf{w}'_1)} \right) + \log \left(\frac{\Phi_2(\mathbf{w}_2)}{\Psi_2(\mathbf{w}_2)} \right) \\ &\quad + \log \left(\frac{\Phi'_2(\mathbf{w}'_2)}{\Psi'_2(\mathbf{w}'_2)} \right) + \log \left(\frac{\Phi_3(\mathbf{w}_3)}{\Psi_3(\mathbf{w}_3)} \right) + \log \left(\frac{\Phi'_3(\mathbf{w}'_3)}{\Psi'_3(\mathbf{w}'_3)} \right), \end{aligned} \quad (27)$$

where $\Phi'_i(\mathbf{w}_i)$, $\Psi'_i(\mathbf{w}_i)$ have similar definition as $\Phi_i(\mathbf{w}_i)$, $\Psi_i(\mathbf{w}_i)$. Hence, we can independently optimize each logarithmic function in (27) which leads to the maximal sum rate. For example, to maximize $\log \left(\frac{\Phi_1(\mathbf{w}_1)}{\Psi_1(\mathbf{w}_1)} \right)$, we formalize the optimization problem similar to (18a)-(18d) as follows

$$\max_{\mathbf{W}_1} \quad \text{trace} \left(\mathbf{h}_1^\perp(1) \mathbf{W}_1 (\mathbf{h}_1^H(1))^\perp \right) \quad (28a)$$

$$\text{s.t.} \quad \text{trace} \left(\mathbf{h}_2^\perp(1) \mathbf{W}_1 (\mathbf{h}_2^H(1))^\perp \right) \leq \frac{\gamma}{4\sigma_h^2} \quad (28b)$$

$$\text{trace}(\mathbf{W}_1) = 1 \quad (28c)$$

$$\mathbf{W}_1 \succeq \mathbf{0} \quad (28d)$$

Rx-1 can minimize the interference leakage power by setting the variables in (25) as given as

$$x_1 = h_{1,2}(4) \mathbf{w}_2 \mathbf{h}_1^H(1) (\mathbf{h}_1(1) \mathbf{h}_1^H(1))^{-1}, \quad (29a)$$

$$x'_1 = h_{1,2}(5) \mathbf{w}_3 \mathbf{h}_1^H(2) (\mathbf{h}_1(2) \mathbf{h}_1^H(2))^{-1}. \quad (29b)$$

The more general form of x_i and x'_i can be found in the subsequence section. Also, because all the logarithmic terms in (27) contain only one beamforming vector, we can optimize each of them independently.

V. ROBUST STOCHASTIC OPTIMIZATION FOR GENERAL K -USER CASES

In this section, we propose a suboptimal scheme to extend the $K = 3$ case of the MAT scheme to general case of K -user, i.e., $\forall K \geq 3$, and achieves $\frac{4K}{3K-1}$ DoF in total. Then we look into the scheme and decompose it into equivalent K e-channels, which can be optimized with the proposed robust optimization method.

A. Achievability scheme

The proposed scheme is inspired by the MAT scheme for the case of $K = 3$. Likewise, in the first phase, linear combinations of private symbols are sent. In the second Phase, the BS reconstructs the interference and sends them to provide another observation for the desired symbol vectors while reducing the interference. Phase 3 is intended to help decoding symbols in Phase 2. The detailed process is listed as follows:

$$\begin{aligned}
 \begin{bmatrix} y_i(1) \\ y_i(2) \\ y_i(3) \\ y_i(4) \\ y_i(5) \\ y_i(6) \end{bmatrix} &= \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{h}_i(1) \\ \mathbf{0} \\ \mathbf{0} \\ h_{i,1}(4)\mathbf{w}_1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{s}_1 + \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{0} \\ \mathbf{h}_i(2) \\ \mathbf{0} \\ \mathbf{0} \\ h_{i,1}(5)\mathbf{w}'_1 \\ \mathbf{0} \end{bmatrix} \mathbf{s}'_1 + \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{h}_i(1) \\ \mathbf{0} \\ \mathbf{0} \\ h_{i,2}(4)\mathbf{w}_2 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{s}_2 \\
 &+ \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{h}_i(3) \\ \mathbf{0} \\ \mathbf{0} \\ h_{i,1}(6)\mathbf{w}'_2 \end{bmatrix} \mathbf{s}'_2 + \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{0} \\ \mathbf{h}_i(2) \\ \mathbf{0} \\ \mathbf{0} \\ h_{i,2}(5)\mathbf{w}_3 \\ \mathbf{0} \end{bmatrix} \mathbf{s}_3 + \sqrt{\frac{P}{2}} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{h}_i(3) \\ \mathbf{0} \\ \mathbf{0} \\ h_{i,2}(6)\mathbf{w}'_3 \end{bmatrix} \mathbf{s}'_3 + \mathbf{n}_i
 \end{aligned} \tag{23}$$

$$\begin{aligned}
 \begin{bmatrix} x_1 y_1(1) - h_{1,2}(4)\mathbf{w}_2 \mathbf{s}_2 \\ y_1(4) - h_{1,2}(4)\mathbf{w}_2 \mathbf{s}_2 \\ x'_1 y_1(2) - h_{1,2}(5)\mathbf{w}_3 \mathbf{s}_3 \\ y_1(5) - h_{1,2}(5)\mathbf{w}_3 \mathbf{s}_3 \end{bmatrix} &= \sqrt{\frac{P}{2}} \begin{bmatrix} x_1 \mathbf{h}_1(1) & 0 \\ h_{1,1}(4)\mathbf{w}_1 & 0 \\ 0 & x'_1 \mathbf{h}_1(2) \\ 0 & h_{1,1}(5)\mathbf{w}'_1 \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}'_1 \end{bmatrix} \\
 &+ \sqrt{\frac{P}{2}} \begin{bmatrix} x_1 \mathbf{h}_1(1) - h_{1,2}(4)\mathbf{w}_2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{s}_2 + \sqrt{\frac{P}{2}} \begin{bmatrix} 0 \\ 0 \\ x'_1 \mathbf{h}_1(2) - h_{1,2}(5)\mathbf{w}_3 \\ 0 \end{bmatrix} \mathbf{s}_3 + \bar{\mathbf{n}}_1
 \end{aligned} \tag{25}$$

Phase 1: There are 4 private symbols intended for each of the K users to be sent in the scheme and will all be sent in each time slot of the first phase. Let us define the 2×1 private symbol vectors for Rx- k as \mathbf{s}_k and \mathbf{s}'_k . In the t th time slot in Phase 1, the BS sends $\mathbf{s}_t + \mathbf{s}'_{t+1}$ and in the K th time slot, $\mathbf{s}_K + \mathbf{s}'_1$ is sent. In total, Phase 1 takes K time slots and at the t th time slot, Rx- k gets $y_k(t) = \mathbf{h}_k(t)\mathbf{s}_t + \mathbf{h}_k(t)\mathbf{s}'_{t+1}$.

Phase 2: This phase also consists of K time slots. Symbol vectors in each of the time slots is constructed according to the symbols received in the corresponding time slot in the first phase. This can be done because all the perfect CSI has been conveyed back to the BS at the end of each time slot. Within the t th time slot of the second phase, i.e., the $(K+t)$ -th time slot in total, the BS reconstructs the symbols as $\mathbf{x}(K+t) = [\mathbf{h}_{t+1}(t)\mathbf{s}_t \quad \mathbf{h}_t(t)\mathbf{s}'_{t+1}]^T$ and we define Rx- $(K+1)$ as Rx-1 to simplify the description. In the t th time slot, there are only symbols for Rx- t and Rx- $(t+1)$. If the two receivers can decode the two symbols in $\mathbf{x}(K+t)$, they can eliminate the interference while obtaining another observation of the desired symbols. For example, Rx- t can eliminate the interference by subtracting $\mathbf{h}_t(t)\mathbf{s}'_{t+1}$ from $y_t(t)$. Meanwhile, Rx- t obtains two observations of \mathbf{s}_t as $\mathbf{h}_{t+1}(t)\mathbf{s}_t$ and $\mathbf{h}_t(t)\mathbf{s}_t$, which enables Rx- t to decode \mathbf{s}_t .

Phase 3: The function of this phase is to help Rx- t and Rx- $(t+1)$ decode $\mathbf{x}(K+t)$, where $t = 1, 2, 3, \dots, K$. Note

that both Rx- t and Rx- $(t+1)$ have already got one linear combination of the two symbols in $\mathbf{x}(K+t)$ as $y_t(K+t)$ and $y_{t+1}(K+t)$, respectively. Therefore, the next step is to provide another linear combination of symbols in $\mathbf{x}(K+t)$ to the two receivers, i.e., $y_i(K+t)$ where $i \neq t, t+1$. To sum up, there are totally K symbols to be sent. We construct the linear combination of the K symbols as given below

$$L_i(y_K(K+1), y_1(K+2), \dots, y_i(K+i+1), \dots, y_{K-1}(2K)), \tag{30}$$

where $i = 1, 2, \dots, K-1$ and $L_i(\cdot)$ means the i th linear combination where the coefficients of each term are shared with the receivers and all the linear combinations are independent. Specifically, each of the linear combinations contains the received signal of each receiver exactly once, which means the receivers have already known one component of the linear combination. Thus, the receiver need $K-1$ independent linear combinations in total to decode them.

At the end of the third phase, the receivers can decode their own desired symbols with backwards decoding. Specifically, after decoding $y_{t-1}(K+t) = h_{t-1,1}(t)\mathbf{h}_{t+1}(t)\mathbf{s}_t + h_{t-1,2}(t)\mathbf{h}_t(t)\mathbf{s}'_{t+1}$ at Rx- t and Rx- $(t+1)$. This is achieved due to the linear combinations in the third phase. Then both receivers will have two observations of linear combinations of $\mathbf{h}_{t+1}(t)\mathbf{s}_t$ and $\mathbf{h}_t(t)\mathbf{s}'_{t+1}$. After decoding them, Rx- t have

knowledge of interference and two observations of the desired signal, which enables it to decoding their own private messages interference-freely.

During the $K + K + (K - 1) = 3K - 1$ time slots of the scheme, there are $4K$ symbols transmitted, therefore the total DoF achieved with K users is $\frac{4K}{3K-1}$, which is lower than the outer bound as $\frac{2K}{K+1}$ [1] but it is higher than the DoF achieved by TDMA scheme.

B. Convex optimization

From the achievability scheme, we can see that Phase 1 is used to transmit new symbols, where no beamforming technique is required. In the second phase, interferences are reconstructed as a linear combination of the desired symbols where the beamforming technique comes in handy. Phase 3 is not an individual phase strictly speaking because it is used to only help in decoding the symbol reconstructed in the second phase. Therefore, similar to the case of $K = 3$, we will decompose the scheme and use the stochastic robust optimization algorithm to find the optimal beamforming vectors in the second phase.

In the second phase, the symbols are constructed as sum of interferences at two relevant receivers at the corresponding time slot. Specifically, at time slot $K + t$, the BS reconstructs the interference generated at Rx- t , Rx- $(t + 1)$ at time slot t , i.e., $\mathbf{h}_{t+1}(t)\mathbf{s}_t$ and $\mathbf{h}_t(t)\mathbf{s}'_{t+1}$. Instead of completely reconstructing interference, we send symbols as $\mathbf{w}_t\mathbf{s}_t$ and $\mathbf{w}'_{t+1}\mathbf{s}'_{t+1}$ to provide flexibility that increases signal detectability. By considering the symmetric assumption, we consider $\mathbf{w}_i, \mathbf{w}'_i$ have unit norm $\forall i \in [1, K]$. It is also to be noted that with the help of Phase 3, both Rx- t , Rx- $(t + 1)$ can successfully acquire $\mathbf{w}_t\mathbf{s}_t$ and $\mathbf{w}'_{t+1}\mathbf{s}'_{t+1}$. Also to the i -th receiver Rx- i , $\mathbf{w}'_{i+1}\mathbf{s}'_{i+1}$ represents approximate interference while $\mathbf{w}_i\mathbf{s}_i$ represents another observation of the desired symbols and the opposite for Rx- $(i + 1)$. By subtracting the approximate interference from the received signal, the effective MIMO channel model for Rx- k to decode symbols \mathbf{s}_k is set up as follows

$$\mathbf{y}_k = \sqrt{\frac{P}{2}} \begin{bmatrix} x_k \mathbf{h}_k(k) \\ h_{k,1}(K+k) \mathbf{w}_k \end{bmatrix} \mathbf{s}_k + \sqrt{\frac{P}{2}} \begin{bmatrix} x_k \mathbf{h}_k(k) - h_{k,2}(K+k) \mathbf{w}'_{k+1} \\ 0 \end{bmatrix} \mathbf{s}'_{k+1} + \bar{\mathbf{n}}_k. \quad (31)$$

Similar to (44), the achievable sum rate is given by

$$I(\mathbf{s}_k; \mathbf{y}_k, \mathbf{w}'_{k+1}\mathbf{s}'_{k+1}) = \mathcal{C} + \log \left(\frac{\Phi_k(\mathbf{w}_k)}{\Psi'_{k+1}(\mathbf{w}'_{k+1})} \right). \quad (32)$$

Therefore, sum of achievable sum rate for all the desired symbols in \mathbf{s}_k and \mathbf{s}'_{k+1} is given as

$$\begin{aligned} I'_\Sigma &= \mathcal{C} + \log \left(\frac{\Phi_k(\mathbf{w}_k)}{\Psi'_{k+1}(\mathbf{w}'_{k+1})} \right) + \log \left(\frac{\Phi'_{k+1}(\mathbf{w}'_{k+1})}{\Psi_k(\mathbf{w}_k)} \right) \\ &= \mathcal{C} + \log \left(\frac{\Phi_k(\mathbf{w}_k)}{\Psi_k(\mathbf{w}_k)} \right) + \log \left(\frac{\Phi'_{k+1}(\mathbf{w}'_{k+1})}{\Psi'_{k+1}(\mathbf{w}'_{k+1})} \right). \end{aligned} \quad (33)$$

Hence, the sum rate can be rewritten as

$$I'_\Sigma = \mathcal{C} + \sum_{k=1}^{K-1} \left[\log \left(\frac{\Phi_k(\mathbf{w}_k)}{\Psi_k(\mathbf{w}_k)} \right) + \log \left(\frac{\Phi'_{k+1}(\mathbf{w}'_{k+1})}{\Psi'_{k+1}(\mathbf{w}'_{k+1})} \right) \right], \quad (34)$$

and then we can maximize each of the logarithmic functions in (34) independently, which only contains one beamforming vector. Therefore, the scheme can be divided into $2K$ e-channels, where the proposed optimization method can maximize the rate to decode \mathbf{s}_k by the following approach,

$$\max_{\mathbf{w}_k} \text{trace} \left(\mathbf{h}_k^\perp(k) \mathbf{W}_k (\mathbf{h}_k^H(k))^\perp \right) \quad (35a)$$

$$\text{s.t.} \quad \text{trace} \left(\mathbf{h}_{k+1}^\perp(k) \mathbf{W}_k (\mathbf{h}_{k+1}^H(k))^\perp \right) \leq \frac{\gamma}{4\sigma_h^2} \quad (35b)$$

$$\text{trace}(\mathbf{W}_k) = 1 \quad (35c)$$

$$\mathbf{W}_k \succeq \mathbf{0} \quad (35d)$$

Similarly, to maximize the achievable sum rate to decode \mathbf{s}'_{k+1} , we have

$$\max_{\mathbf{w}'_{k+1}} \text{trace} \left(\mathbf{h}_{k+1}^\perp(k) \mathbf{W}'_{k+1} (\mathbf{h}_{k+1}^H(k))^\perp \right) \quad (36a)$$

$$\text{s.t.} \quad \text{trace} \left(\mathbf{h}_k^\perp(k) \mathbf{W}'_{k+1} (\mathbf{h}_k^H(k))^\perp \right) \leq \frac{\gamma}{4\sigma_h^2} \quad (36b)$$

$$\text{trace}(\mathbf{W}'_{k+1}) = 1 \quad (36c)$$

$$\mathbf{W}'_{k+1} \succeq \mathbf{0} \quad (36d)$$

In this way, with different values of k , the optimal beamforming vectors are found and thus the achievable sum rates are maximized. When dealing with the received signals, receivers firstly construct e-channel for the desired symbol, such as (31). Then receivers select the variables in the receive signal matrix as given by

$$x_k = h_{k,2}(K+k) \mathbf{w}'_{k+1} \mathbf{h}_k^H(k) (\mathbf{h}_k(k) \mathbf{h}_k^H(k))^{-1}, \quad (37a)$$

$$x'_{k+1} = h_{k,2}(K+k) \mathbf{w}_k \mathbf{h}_{k+1}^H(k) (\mathbf{h}_{k+1}(k) \mathbf{h}_{k+1}^H(k))^{-1}. \quad (37b)$$

VI. SIMULATION RESULTS AND DISCUSSION

In order to realize the full potential of the robust stochastic optimization method, channel correlation is considered in the simulations where the tradeoff between enhancing signal detectability and aligning interference is studied. The correlated channel can be modeled as follows

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2}, \quad (38)$$

where \mathbf{R}_r and \mathbf{R}_t are receiver and transmitter correlation matrices with diagonal elements being *one* and others being r_t and r_r , $r_t, r_r \in [0, 1]$, respectively [21], [22]. Matrix \mathbf{H}_w denotes the channel with i.i.d. entries. To more clearly show the performance gain, relatively large channel correlation is assumed, i.e., $r_t = 0.7$ and $r_r = 0.6$. The dual-SINR scheme is simulated by taking the closed-form solution of the beamforming vectors in [11] into the effective MIMO channel. Notice that each row of \mathbf{H} is regarded as the channel vector from BS to the receivers. Each point of the rest of the simulations are conducted through 200 channel realizations.

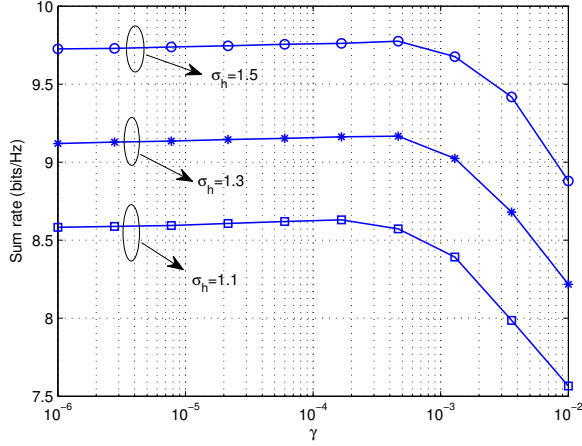


Fig. 1. Sum rate for 2-user BS versus γ with various σ_h .

Figure 1 illustrates the sum-rate versus γ for various channel variances σ_h . The transmitting power is $P = 20$ dB with various channel variances, i.e., $\sigma_h = 1.5$, $\sigma_h = 1.3$ and $\sigma_h = 1.1$. This result enables us to select the optimal γ . There is no need for the transmitters to conduct this simulation every time when designing the precoding vector. Instead, γ curve only needs to be simulated once and stored in the memory of transmitters. When transmitters acquired the channel variance, it can look up the stored values and find the proper γ . When $\gamma \rightarrow 0$, (18b) makes the interference almost perfectly aligned, which reduces to the MAT scheme (according to the numerical simulation, the MAT scheme witnesses an interference leakage power at the level of $\gamma = 10^{-32}$). When γ increases, more weights of the beamforming vectors are used to increase the linear independence of the two signal data streams. It can be observed that above certain threshold, the increasing interference leakage power starts to offset the increasing detectability of the intended signal. The optimal γ is the value that keeps the balance between aligning interference and increasing the linear independence between two observations of the desired signals. It is clear from this figure that the optimal γ is smaller compared to the channel variance. Therefore, the interference power can be regarded as reduced at a high level. In other words, aligning the interference vectors is more important than increasing the detectability of the desired symbols. This is the reason that the proposed scheme does not outperform the MAT scheme considerably.

Robustness of the proposed scheme can be seen in Figure 2 where results from the MAT scheme, dual SINR scheme [11] and the proposed robust stochastic optimization method are depicted for comparison. Sum rates, defined in (39), are drawn with respect to signal-to-noise ratio (SNR) and $K = 2$ users. In this figure, channel is subjected to little volatility, i.e., channel variance $\sigma_h = 0.2$ and γ is chosen to be 4×10^{-6} , which is chosen through numerical approach similar to the first simulation. We use the same numerical simulation to choose γ in the remaining simulations. From this figure, it can be observed that the proposed scheme and dual SINR scheme almost overlap with each other in low SNR range. With the increase of SNR, the proposed scheme outperforms both of

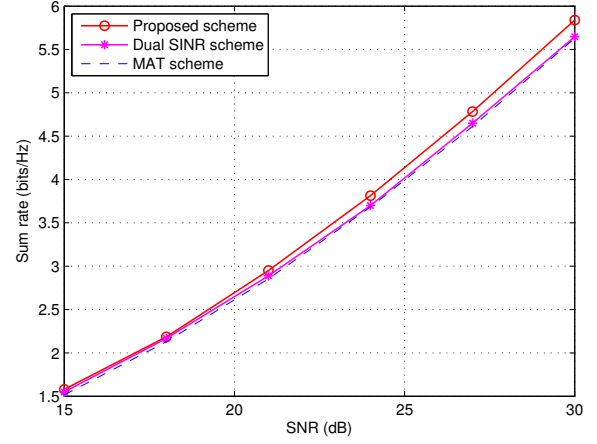


Fig. 2. Sum rate of 2-user BC with channel variance $\sigma_h = 0.2$

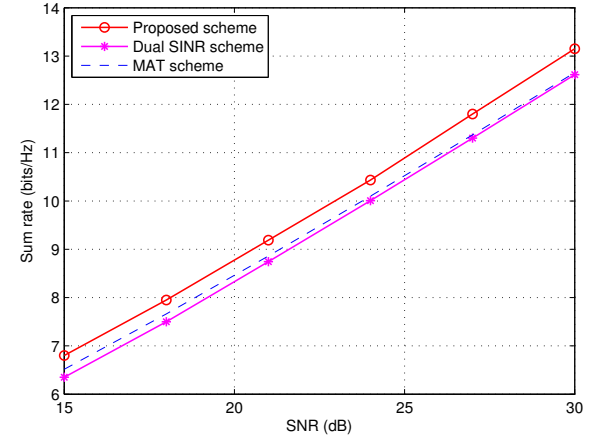


Fig. 3. Sum rate of 2-user BC with channel variance $\sigma_h = 1.5$.

the dual SINR scheme and the MAT scheme.

Similar to the first simulation, the results shown in Figure 3 depict the sum rate versus SNR, except that channel variance is assumed to be $\sigma_h = 1.5$ and γ is selected to be 2×10^{-4} . The proposed scheme outperforms both dual SINR scheme and the MAT scheme with channel entries differing from standard normal distribution. Simulations in Figure 2 and Figure 3 both validate the robustness brought by the probabilistic constraint in various channel variances.

A comparison between the proposed scheme when $K = 3$ and $K = 4$, MAT scheme and TDMA is depicted in Figure 4, where $\sigma_h = 1.5$ and $\gamma = 2 \times 10^{-4}$. Because optimal scheme has been given in $K = 3$ case, where 1.5 DoF are available, it outperforms the $K = 4$ case, where only suboptimal scheme is given and 1.45 DoF are available. Although non-optimal, the scheme for the case of $K = 4$ still makes use of the delayed CSIT and considerably outperforms the TDMA scheme, where no delayed CSIT is utilized and one DoF is achieved. Meanwhile, the curve with $K = 4$ outperforms the MAT scheme with 1.5 DoF, due to the robustness of the proposed scheme. High channel correlation makes the TDMA scheme outperform the MAT scheme in the low SNR region.

Figure 5 illustrates the achievable sum rate with different γ . In this simulation, $\sigma_h = 1.5$. It can be seen that when $\gamma = 2 \times$

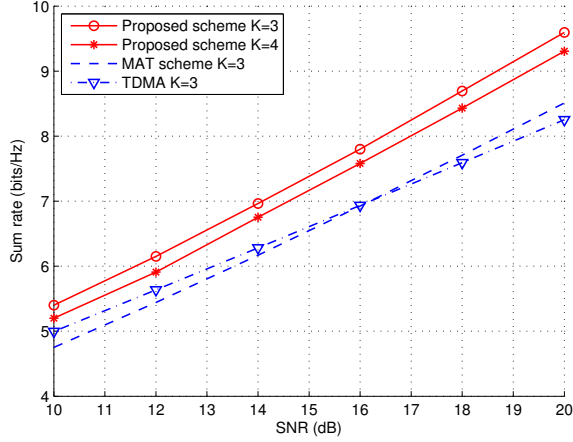


Fig. 4. Sum rate of 3- and 4-user BC compared with the MAT and TDMA scheme.

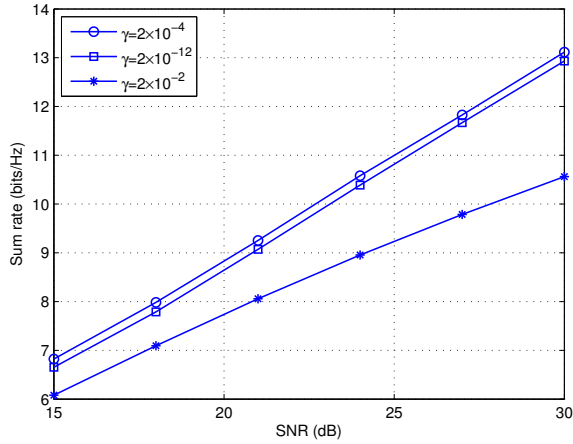


Fig. 5. Sum rate of 2-user BC compared with different γ .

10^{-2} , the interference leakage power is dominant and severely reduces the sum rate. When $\gamma = 2 \times 10^{-12}$, strict interference leakage power constraint makes the proposed scheme reducing to the MAT scheme which gives too little flexibility to the precoding vectors and interference suppression matrices.

VII. CONCLUSION

In this work, we proposed an iterative-free robust stochastic optimization method to enhance the achievable sum rate for K -user MISO BC using delayed CSIT where the BS was only equipped with 2 transmitting antennas. The CSI was perfectly estimated by the receivers and conveyed back to the BS without loss of any accuracy at the end of each time slot. The proposed technique develops by generalizing the beamforming vectors and the interference suppression matrices in the 2-user MAT scheme, which endowed both of them with more flexibility. By exploiting the distribution of the channel entries, we formalized a probabilistic constraint which constrained the interference leakage power while bringing robustness to the scheme. We also proposed a suboptimal achievability scheme (optimal when $K = 3$) for general K -user case which achieved $\frac{4K}{3K-1}$ DoF. We carefully researched the 3-user case as well as the suboptimal general K -user case and inferred

that both cases could be divided into multiple effective MIMO channel similar to the effective MIMO channel when $K = 2$. The latter property enabled the proposed robust stochastic optimization method to maximize the achievable sum rate for general K -user case. The simulation results showed that the proposed scheme outperformed the MAT scheme and the dual SINR scheme with various channel variances. In addition, the tradeoff between aligning interference (decreasing interference leakage power) and increasing the linear independence of two signal data streams (increasing detectability of signal) was validated by simulations, which showed that the achievable sum rate increased slowly with the allowed interference leakage power and then decreased drastically. This explains why neither the proposed work or dual SINR scheme could not outperform the MAT scheme considerably.

APPENDIX A PROOF OF THEOREM 1

Start from the definition of achievable sum rate, which for the e -channel should written as

$$I(\mathbf{s}_i; \mathbf{y}_i) = \log \det \left(\mathbf{I} + \frac{P}{2} \mathbf{W}_i^{-1} \underbrace{\left(\mathbf{I} + \frac{P}{2} \mathbf{H}_{ij} \mathbf{H}_{ij}^H \right)^{-1} \mathbf{H}_{ii} \mathbf{H}_{ii}^H}_{\mathbf{Q}_i} \right), \quad (39)$$

where $i \neq j$, \mathbf{I} is the 2×2 identity matrix and $\mathbf{W}_i = \text{diag}\{1, \max\{1, |x_i|^2\}\}$ is the expectation of covariance matrix of the white noise. To simplify the following description, denote $\max\{1, |x_i|^2\}$ as \hat{x}_i^2 . Again, let us consider the receiver Rx-1 as an example. With \mathbf{H}_{11} and \mathbf{H}_{12} defined in (5), we obtain the following

$$\mathbf{Q}_1 = \begin{bmatrix} \|\mathbf{h}_1(1)\|^2 & h_{1,1}^*(3) \mathbf{h}_1(1) \mathbf{w}_1^H \\ \frac{h_{1,1}(3) \mathbf{w}_1 \mathbf{h}_1^H(1)}{\Theta_1} & \frac{|h_{1,1}(3)|^2 \|\mathbf{w}_1\|^2}{\Theta_1} \end{bmatrix}, \quad (40)$$

where

$$\Theta_1 = 1 + \frac{P}{2} \|\mathbf{h}_{1,1}(3) \mathbf{w}_2 - x_1 \mathbf{h}_1(2)\|^2. \quad (41)$$

The equality inside the logarithmic function at Rx-1 can be expanded as

$$\begin{aligned} & \det \left(\mathbf{I} + \frac{P}{2} \mathbf{W}_1^{-1} \mathbf{Q}_1 \right) \\ &= \det \left(\begin{bmatrix} 1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2 & \frac{P}{2} h_{1,1}^*(3) \mathbf{h}_1(1) \mathbf{w}_1^H \\ \frac{P h_{1,1}(3) \mathbf{w}_1 \mathbf{h}_1^H(1)}{2\Theta_1 \hat{x}_1^2} & 1 + \frac{P |h_{1,1}(3)|^2 \|\mathbf{w}_1\|^2}{2\Theta_1 \hat{x}_1^2} \end{bmatrix} \right) \\ &= 1 + \frac{P}{2} \left(\|\mathbf{h}_1(1)\|^2 + \frac{|h_{A1}(3)|^2 \|\mathbf{w}_1\|^2}{\Theta_1 \hat{x}_1^2} \right) \\ & \quad + \frac{P^2}{4} \left(\frac{|h_{A1}(3)|^2 \|\mathbf{h}_1(1)\|^2 \|\mathbf{w}_1\|^2 - |h_{A1}(3)|^2 \mathbf{w}_1 \mathbf{h}_1(1) \mathbf{h}_1^H(1) \mathbf{w}_1^H}{\Theta_1 \hat{x}_1^2} \right) \end{aligned} \quad (42)$$

Since logarithm function is monotonically increasing, then maximization of the achievable sum rate is equivalent to maximization of (42).

Because the transmission power is limited to P , we have the power limitation of the beamforming vectors as $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \leq 2$. In order to achieve the maximal sum rate, the equality should always be valid, i.e., $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 = 2$. Since we assume all the channel entries are i.i.d., then we

$$I(\mathbf{s}_1; \mathbf{y}_1) = \log \left(1 + \frac{P}{2} \left(\|\mathbf{h}_1(1)\|^2 + \frac{|h_{1,1}(3)|^2}{\Theta_1 \hat{x}_1^2} \right) \frac{P^2}{4} \left(\frac{|h_{1,1}(3)|^2 \|\mathbf{h}_1(1)\|^2 - |h_{1,1}(3)|^2 \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H}{\Theta_1 \hat{x}_1^2} \right) \right) \quad (43)$$

$$= \log \left(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2 \right) + \log \left(1 + \frac{\frac{P}{2} |h_{1,1}(3)|^2 + \frac{P^2}{4} |h_{1,1}(3)|^2 (\|\mathbf{h}_1(1)\|^2 - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H)}{(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2) \Theta_1 \hat{x}_1^2} \right) \quad (44)$$

$$I(\mathbf{s}_2; \mathbf{y}_2) = \log \left(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2 \right) + \log \left(1 + \frac{\frac{P}{2} |h_{2,1}(3)|^2 + \frac{P^2}{4} |h_{2,1}(3)|^2 (\|\mathbf{h}_2(2)\|^2 - \mathbf{w}_2 \mathbf{h}_2^H(2) \mathbf{h}_2(2) \mathbf{w}_2^H)}{(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2) \Theta_2 \hat{x}_2^2} \right), \quad (45)$$

$$I(\mathbf{s}_1; \mathbf{y}_1) + I(\mathbf{s}_2; \mathbf{y}_2) \approx \log \left(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2 \right) + \log \left(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2 \right) \quad (46a)$$

$$+ \log \left(\frac{\frac{P}{2} |h_{1,1}(3)|^2 + \frac{P^2}{4} |h_{1,1}(3)|^2 (\|\mathbf{h}_1(1)\|^2 - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H)}{(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2) \Theta_1 \hat{x}_1^2} \right) \quad (46b)$$

$$+ \log \left(\frac{\frac{P}{2} |h_{2,1}(3)|^2 + \frac{P^2}{4} |h_{2,1}(3)|^2 (\|\mathbf{h}_2(2)\|^2 - \mathbf{w}_2 \mathbf{h}_2^H(2) \mathbf{h}_2(2) \mathbf{w}_2^H)}{(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2) \Theta_2 \hat{x}_2^2} \right) \quad (46c)$$

$$= \log \left(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2 \right) + \log \left(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2 \right) \quad (46d)$$

$$+ \log \left(\frac{\frac{P}{2} |h_{1,1}(3)|^2 + \frac{P^2}{4} |h_{1,1}(3)|^2 (\|\mathbf{h}_1(1)\|^2 - \mathbf{w}_1 \mathbf{h}_1^H(1) \mathbf{h}_1(1) \mathbf{w}_1^H)}{(1 + \frac{P}{2} \|\mathbf{h}_1(1)\|^2) \Theta_2 \hat{x}_2^2} \right) \quad (46e)$$

$$+ \log \left(\frac{\frac{P}{2} |h_{2,1}(3)|^2 + \frac{P^2}{4} |h_{2,1}(3)|^2 (\|\mathbf{h}_2(2)\|^2 - \mathbf{w}_2 \mathbf{h}_2^H(2) \mathbf{h}_2(2) \mathbf{w}_2^H)}{(1 + \frac{P}{2} \|\mathbf{h}_2(2)\|^2) \Theta_1 \hat{x}_1^2} \right). \quad (46f)$$

could reasonably implement symmetric power allocation to approximate the realistic situation, i.e., $\|\mathbf{w}_1\| = \|\mathbf{w}_2\| = 1$, which can simplify the achievable sum rate at Rx-1 and Rx-2 as in (44) and (45), respectively, where $\Theta_2 = 1 + \frac{P}{2} \|\mathbf{h}_{2,1}(3)\mathbf{w}_1 - x_2 \mathbf{h}_2(1)\|^2$. Hence, the total achievable sum rate is given by (46a)-(46f). Note that (46a) is valid in high SNR cases and equation (46d) is because of the following $\log \left(\frac{A}{B} \right) + \log \left(\frac{C}{D} \right) = \log \left(\frac{AC}{BD} \right) = \log \left(\frac{A}{D} \right) + \log \left(\frac{C}{B} \right)$.

APPENDIX B PROOF OF THEOREM 2

By rearranging (14b), we obtain the following

$$\Pr \left(|h_{2,1}(3)|^2 \leq \frac{\alpha}{\text{trace}(\mathbf{w}_1 (\mathbf{h}_2^H(1))^\perp \mathbf{h}_2^\perp(1) \mathbf{w}_1^H)} \right) \geq \beta \quad (47)$$

Let $\Theta \triangleq \frac{\alpha}{\text{trace}(\mathbf{w}_1 (\mathbf{h}_2^H(1))^\perp \mathbf{h}_2^\perp(1) \mathbf{w}_1^H)}$, and then the previous expression can be simplified into following equation

$$\Pr \left(|h_{2,1}(3)| \leq \sqrt{\Theta} \right) \geq \beta. \quad (48)$$

In this inequality, we have knowledge of all the parameters except for $h_{2,1}(3)$ and with the help of its distribution, the constraint becomes to find the cumulative distribution function (CDF) of $|h_{A1}(3)|$. Typically, we assume the real and imaginary part of the channel are i.i.d., that is we have

$$\text{Re}\{h_{A1}(3)\} \sim \mathcal{N}(0, \sigma_h^2); \quad \text{Im}\{h_{A1}(3)\} \sim \mathcal{N}(0, \sigma_h^2) \quad (49)$$

where $\text{Re}(\bullet)$ and $\text{Im}(\bullet)$ are the real and imaginary parts respectively and σ_h^2 is the variance of channel entries with urban practical value 2 dB \sim 4 dB [23]. Using the latter property together with the probabilistic constraint, we can approximate the constraint into two parts as follows

$$\begin{aligned} & \Pr \left(|h_{2,1}(3)| \leq \sqrt{\Theta} \right) \\ & \geq \Pr \left(|\text{Re}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \cap \Pr \left(|\text{Im}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \\ & = \Pr \left(|\text{Re}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \cdot \Pr \left(|\text{Im}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \\ & = \left[\Pr \left(|\text{Re}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \right]^2 \end{aligned} \quad (50)$$

With (48) and (50), we can further write the probabilistic constraint as

$$\Pr \left(|\text{Re}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \geq \sqrt{\beta}, \quad (51)$$

or equivalently

$$\Pr \left(|\text{Im}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \geq \sqrt{\beta}, \quad (52)$$

where we consider the real part in detail. Because the real part of the channel entries is Gaussian distributed with zero mean and σ_h^2 variance, we can further rewrite the real part of the

probabilistic constraint as

$$\begin{aligned} & \Pr \left(|\operatorname{Re}\{h_{2,1}(3)\}| \leq \sqrt{\frac{\Theta}{2}} \right) \\ &= \Pr \left(\operatorname{Re}\{h_{2,1}(3)\} \leq \sqrt{\frac{\Theta}{2}} \right) - \Pr \left(\operatorname{Re}\{h_{2,1}(3)\} \leq -\sqrt{\frac{\Theta}{2}} \right) \\ &= \frac{1}{2} \left[\operatorname{erf} \left(\frac{\sqrt{\Theta}}{2\sigma_h} \right) - \operatorname{erf} \left(\frac{-\sqrt{\Theta}}{2\sigma_h} \right) \right], \end{aligned} \quad (53)$$

where the CDF function of Gaussian distribution, i.e.,

$$F(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{x}{\sqrt{2}\sigma_h} \right), \quad (54)$$

and $\operatorname{erf}(x)$ is the standard error function for Gaussian distribution. After achieving (53), we can further rewrite the probabilistic constraint as

$$\operatorname{erf} \left(\frac{\sqrt{\Theta}}{2\sigma_h} \right) \geq \sqrt{\beta} \Rightarrow \Theta \geq 4\sigma_h^2 \left[\operatorname{erf}^{-1}(\sqrt{\beta}) \right]^2. \quad (55)$$

and then by expanding the inequality, we obtain the following

$$\operatorname{trace} \left(\mathbf{w}_1 (\mathbf{h}_2^H(1))^\perp \mathbf{h}_2^\perp(1) \mathbf{w}_1^H \right) \leq \frac{\alpha}{4\sigma_h^2 \left[\operatorname{erf}^{-1}(\sqrt{\beta}) \right]^2}. \quad (56)$$

Because the real and imaginary part of the channel entries are i.i.d., we should obtain exactly the same constraint if we use the imaginary part of (50), which can be neglected without any loss of any generality.

REFERENCES

- [1] M. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, 2012.
- [2] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K -user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, Aug. 2008.
- [3] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.
- [4] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, 2013.
- [5] X. Yi, S. Yang, D. Gesbert, and M. Kobayashi, "The degrees of freedom region of temporally-correlated MIMO networks with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, Sep. 2013.
- [6] J. Chen and P. Elia, "Degrees-of-freedom region of the MISO broadcast channel with general mixed-CSIT," *ArXiv e-prints*, May 2012.
- [7] J. Chen, S. Yang, and P. Elia, "On the fundamental feedback-vs-performance tradeoff over the MISO-BC with imperfect and delayed CSIT," in *IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013.
- [8] J. Chen and P. Elia, "Can imperfect delayed CSIT be as useful as perfect delayed CSIT? DoF analysis and constructions for the BC," in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 1254–1261.
- [9] Z. Wang, M. Xiao, C. Wang, and M. Skoglund, "Degrees of freedom of multi-hop MIMO broadcast networks with delayed CSIT," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 1–4, 2013.
- [10] S. Yang, M. Kobayashi, P. Piantanida, and S. Shamai, "Secrecy degrees of freedom of MIMO broadcast channels with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5244–5256, 2013.
- [11] X. Yi and D. Gesbert, "Precoding on the broadcast MIMO channel with delayed CSIT: The finite SNR case," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 2933–2936.
- [12] —, "Precoding methods for the MISO broadcast channel with delayed CSIT," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1–11, 2013.
- [13] Y. Lejosne, D. Slock, and Y. Yuan-Wu, "Degrees of freedom in the MISO BC with delayed-CSIT and finite coherence time: Optimization of the number of users," in *International Conference on Network Games, Control and Optimization (NetGCoP)*, 2012, pp. 80–85.
- [14] S. Vorobyov, A. Gershman, and Y. Rong, "On the relationship between the worst-case optimization-based and probability-constrained approaches to robust adaptive beamforming," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, vol. 2, 2007.
- [15] Y. Rong, S. Vorobyov, and A. Gershman, "Robust linear receivers for multiaccess space-time block-coded MIMO systems: a probabilistically constrained approach," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1560–1570, 2006.
- [16] S. Vorobyov, A. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, 2003.
- [17] A. J. G. Anandkumar, A. Anandkumar, S. Lambotharan, and J. Chambers, "Robust rate maximization game under bounded channel uncertainty," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4471–4486, 2011.
- [18] J. Wang, M. Bengtsson, B. Ottersten, and D. Palomar, "Robust MIMO precoding for several classes of channel uncertainty," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3056–3070, 2013.
- [19] S. Vorobyov, Y. Rong, and A. Gershman, "Robust adaptive beamforming using probability-constrained optimization," in *IEEE/SP 13th Workshop on Statistical Signal Processing*, 2005, pp. 934–939.
- [20] A. Wiesel, Y. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4409–4418, Sep. 2008.
- [21] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.
- [22] C. Xiao, J. Wu, S. Leong, Y. Zheng, and K. Letaief, "A discrete-time model for triply selective MIMO rayleigh fading channels," *IEEE Trans. Wireless Commun.*, 2004.
- [23] G. Turin, F. Clapp, T. Johnston, S. Fine, and D. Lavry, "A statistical model of urban multipath propagation," *IEEE Trans. Veh. Technol.*, vol. 21, no. 1, pp. 1–9, 1972.



Yi Luo (S'11) received the B.Eng. degrees from University of Electronic Science and Technology of China, Chengdu, China, in 2010, in Solid-State Electronics. Since 2011, he is pursuing the Ph.D. degree at Institute of Digital Communication, The University of Edinburgh, Edinburgh, United Kingdom. His current research interests include stochastic geometry-based analysis for wireless communication and interference alignment.



Tharmalingam Ratnarajah (A'96-M'05-SM'05) received the B.Eng. (Hons.), M.Sc., and Ph.D. degrees. He is currently with the Institute for Digital Communications, the University of Edinburgh, Edinburgh, U.K., as a Reader in signal processing and communications. His research interests include random matrices theory, information theoretic and signal processing aspects of wireless networks, statistical and array signal processing, biomedical signal processing, and quantum information theory. He has published over 250 publications in these

areas and holds four U.S. patents. He is currently the coordinator of the FP7 projects HARP (3.2M €) in the area of highly distributed MIMO and ADEL (3.7M €) in the area of dynamic spectrum access. Previously, he was the coordinator of FP7 Future and Emerging Technologies project CROWN (2.3M €) in the area of cognitive radio networks and HIATUS (2.7M €) in the area of interference alignment. Dr. Ratnarajah is a member of the American Mathematical Society and Information Theory Society and Fellow of Higher Education Academy (FHEA), UK.